ARCHIVING GIS-TYPE PRODUCTS IN THE PDS. L. R. Gaddis and T. M. Hare, USGS Astrogeology Science Center, Flagstaff, AZ 86001 (lgaddis@usgs.gov)

Introduction: In 2015, NASA tasked the Planetary Data System (PDS) with supporting archival products from science investigators funded by several NASA data analysis and research programs. This opened the door to the need for archival versions of numerous product types previously unsupported by PDS, including data formats and structures that are proprietary (e.g., Geographic Information System or GIS vector files such as Esri geodatabases and shapefiles). At nearly the same time (starting in 2013), efforts were underway by the PDS to support development of archives under the new Planetary Data System Version 4 (PDS4) data standard [1-3]. Here we address how the Cartography and Imaging Sciences (Imaging, or IMG) discipline node of the PDS is working to support these NASA requirements.

Data Formats in PDS: The PDS describes acceptable archival file data formats in the PDS4 Information Model [IM; 4; see https://pds.nasa.gov/datastandards/documents/im/); these also are summarized in the PDS4 Data Provider's Handbook [5]. The PDS4 IM is a set of well-defined concepts, objects, relationships, rules and operations that represent the contents and structure of PDS archives. The IM defines archival data formats and drives the development of PDS4 documentation and tools. The IM is revised as the needs of data providers and users evolve (usually once a year), as is the Data Provider's Handbook.

Four basic structural data formats are allowed in PDS4 (see p. 12 of [5]):

1) ***Arrays***, suitable for raster images with two or more (up to 16) dimensions. The elements of an array are binary and homogeneous, and all elements must have the same data type. The individual elements of any array are stored with their bytes in the order dictated by their scalar (i.e., least- or most-significant byte first, specified by the provider). Array elements are stored in the axis order of last index varying fastest (see Section 4A of [6]). If formatted appropriately, file formats including FITS 3.0, ISIS3, VICAR1/2, and GeoTiff can be used to store the raster image.

2) ***Repeating record structures***, suitable for tables with fixed-width columns. The data may be either binary or character, but not both. The fields of a record may be heterogeneous – they may have different data types within the binary or character data. Any single field must be homogeneous from one record to another.

3) ***Parsable structures***, suitable for plain text and for tables with variable-length fields and records (i.e., delimited text such as comma-separated value or CSV format). The contents are a byte stream that can be parsed with standard rules (e.g., comma separated entries, standard punctuation); no decoding software is required. Examples are 7-Bit ASCII Text and UTF-8 Text, etc.

4) ***Encoded structures***, suitable for documents, browse products, etc., but generally not for science data products. Contents are a byte stream decoded by software before use (e.g., with Adobe Acrobat©). The use of encoded structure objects is restricted by PDS to a limited set of PDS-approved external standards (e.g., PDF/A, JPEG, and GIF). Only in exceptional cases are encoded structure objects allowed for storing observational (science) data. Prior PDS approval is required, and the process may take several months.

The PDS4 IM [4] defines a base class for each of these four structures. A base class is a generic definition that is built upon to form more specific definitions, without losing the basic features of the class. For example, the Array base class is used to form the more specific Array_2D class and from that the

Array_2D_Image class. While these four structures cover all types of products in a PDS archive, there are further restrictions on observational (science) data. The philosophy of PDS is that to preserve data for the long term, formats must be as simple as possible, well described, and not reliant on specific software, because that software may be unavailable in the future. To help make observational data more easily available to the science community, PDS allows copies of data to be archived in *supplemental* formats (see https://pds.nasa.gov/datastandards/documents/policy/Supplemental_Formats.pdf). Examples are MPEG-4, JPEG2000, SEED 2.4, etc. PDS provides software to transform data from archival to more popular formats (https://pds.nasa.gov/tools/about/transform/).

GIS-type Products in PDS: ***Proprietary file formats such as Esri shapefiles and geodatabases are not PDS4-compliant products, and they cannot be archived in their native formats in PDS.*** However, geodatabase components (layers) can typically be archived separately in PDS4-compliant formats as tables or image arrays. For archiving GIS vector data (shapefiles), the PDS has approved a variant of the GeoCSV format [7]. GeoCSV is a CSV table in which the last field maintains a Well-Known Text geometry string to describe GIS points, lines or polygons. Data providers who wish to archive such products with the PDS should be prepared to work with the PDS discipline nodes to isolate, translate, and/or reformat their GIS-type data products as needed.

The Annex: Smaller research products from individual, NASA-funded science investigators needing to be archived can be delivered to IMG and served via the ***Annex*** (https://astrogeology.usgs.gov/pds/annex), a facility hosted at the U.S. Geological Survey's Astrogeology Science Center. The Annex was developed by IMG to support scientists who use PDS data to create derived geospatial products that can be registered to a solid planetary body [8]. Since 2016, the Annex has been considered a PDS-equivalent site for hosting such data products. As of 2015, PDS products delivered to IMG for serving via the Annex will be archived in PDS4 formats and subject to review. Providers delivering products (e.g., figures, tables) associated with publications have the option of PDS4 reformatting and review.

On each product page in the Annex, detailed metadata are included. Data providers must provide metadata for each product, including originator name and contact information, geographic coordinates, target body, descriptive caption, publication date, lineage and data source, validation and review status, quality and completeness assessments, linkages to other PDS products, and citations. The metadata standard used for Astropedia was created by the U.S. Federal Geographic Data Committee (FGDC) with small modifications to better support planetary data [9, 10]. These detailed metadata support development of the PDS4 archives and improved searches.

References: [1] Crichton et al. (2011) EPSC Abstracts, 6, #1733. [2] Beebe et al. (2010) AAS-DPS meeting #42, id.37.02; Bulletin of the American Astronomical Society, Vol. 42, p. 967. [3] Neakrase et al. (2012) 3rd Int. Planet. Dunes Wkshp, abs 7049. [4] Hughes et al. (2018) Planetary & Space Science 150, pp. 43-49. [5] The PDS4 Data Provider's Handbook, https://pds.nasa.gov/datastandards/documents/dph/current/PDS4_DataProvidersHandbook_1.11.0.pdf). [6] Planetary Data System Standards Reference, 1.9.0 (https://pds.nasa.gov/pds4/doc/sr/current/). [7] Hare (2018), http://astropedia.astrogeology.usgs.gov/download/Docs/Workshops/PDSCouncil/Meeting2018/Hare_GIS4PDS4.pptx. [8] Hare et al., this meeting. [9] Federal Geographic Data Committee (2011) Preparing for International Metadata, Federal Geographic Data Committee, Washington, D.C.; http://www.fgdc.gov/. [10] Hare et al. (2011) LPS 42, #2154.